Genomic organization of the retinoic acid receptor gamma gene

Jürgen M.Lehmann, Birgit Hoffmann and Magnus Pfahl*
Cancer Research Center, La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

Received October 9, 1990; Revised and Accepted December 19, 1990 Genbank accession no. M38258

ABSTRACT

The retinoic acid receptors (RAR) belong to the large family of ligand responsive gene regulatory proteins that includes receptors for steroid and thyroid hormones. These proteins contain two highly conserved domains, involved in determining their DNA and ligand binding activities. Three distinct RARs have been identified (RARα, β, and γ) which are encoded by genes on separate chromosomes. Additional isoforms of the three receptors have been described that all differ in the N-terminal regions. To gain insight into the genomic organization and mechanisms of RAR isoform generation, we have analyzed the genomic structure of the RARγ gene. The major portion of the RARγ protein, including DNA and ligand binding domains, is encoded by seven exons that are identical for all RARg isoforms and are represented by a relatively small portion of the RARγ gene. The major portion of this gene encodes separate N-terminal exons for γ1 and γ2 isoforms and several exons for γ1 untranslated regions. We show that RARγ2 transcription is regulated by its own promoter. In comparison with the steroid receptor subfamily, various splice sites of RARγ occur at altered positions, suggesting that the RAR subfamily has diverged early during evolution.

INTRODUCTION

Retinoic acid (RA) is a vitamin A metabolite, which exerts essential effects in the control of epithelial cell growth, fetal development, cellular differentiation (1), and has recently been identified as a natural morphogen conferring pattern formation of the regenerating amphibian and chick limb bud (2,3). A potential model for the molecular mechanism underlying these pleiotropic effects was provided by the identification of RA receptors (RARs) as members of the steroid/thyroid hormone receptor superfamily (4–7). These are ligand-activated transcription factors which interact specifically with cognate DNA elements (i.e. hormone response elements) to modulate transcription (reviewed in refs. 8,9). They all contain a highly conserved DNA-binding domain of the 'zinc-finger' type and a predominantly hydrophobic ligand-binding domain that also encodes the receptor dimerization function (10–12).

The isolation of three different human RAR subtypes: RARα (4,5), RARβ (6, 7), and RARγ (13, 14) suggests that these have different functions and may activate distinct target genes. The three RAR genes reveal a spatially and temporally restricted expression pattern during mouse embryogenesis. The RARγ expression pattern suggests a crucial role in the regulation of gene expression during morphogenesis and differentiation of squamous epithelia (15, 16). Our recent finding that RARβ and RARγ show distinct activation profiles of the two different RA response elements,TRE-pal and βRARE, indicates for the first time the existence of RAR target gene specificity (17). Since both RARs appear to bind equally to both response elements, their differential activation capacity might be attributed to their different N-terminal regions. This region is highly divergent among RARs but is highly conserved between different species for the same RAR isoform (18).

Recently, a second human RARγ, designated as the RARγ2, was isolated. This differs from RARγ1 (former RARγ) in the N-terminal region but is otherwise identical (13,14,18). The thyroid hormone receptors (TRs) exist also as variants that differ either in their amino- or carboxy-terminal ends (19–21 and references therein). TR-α splice variants with altered C-terminal portions, are not transcriptional activators but acts as hormone-independent repressors (19, Herrmann et al., submitted). The delineation of mechanisms that generate expression of functionally distinct RAR isoforms will provide insight into the highly specific network that allows the single substance, RA, to have cell lineage specific regulatory functions.

To determine the mechanisms regulating the synthesis of the different RARγ isoforms and to show how the exon structure contributes to the functional domains, the genomic organization of the RARγ gene was determined. The 3'-third of the gene was found to contain almost all coding exons except for the N-terminal A- region exon. The exons encoding the alternative A-regions of RARγ1 and RARγ2 are contained in the first two-thirds of the RARγ gene. We show that RARγ2 transcription is regulated by its own promoter. Comparison of the RARγ exon structures with those of steroid hormone receptors suggest that the RAR

* To whom correspondence should be addressed

subfamily developed independently from the steroid receptor subfamily.

MATERIALS AND METHODS

Isolation and sequencing of human RARγ cDNA and genomic clones

The RARγ1 cDNA clone was obtained from screening a human placental λgt11 expression library (22) kindly provided by J.L. Millan from this institute. A random primed ^{32}P-labeled cDNA comprising the ligand-binding domain of RARα was used as a hybridizing probe. This cDNA fragment was obtained by using the PCR-technique and phage DNA from the same placental library (23). Standard hybridization and stringent washing conditions were used (24). By using the 5′ AccI fragment derived from cDNA clone RARγ1 (position 1−403) as a probe, two clones designated as $RAR\gamma_{gene1-2}$ were isolated from a human genomic placental library (Stratagene). Genomic walking using 5′-sequences of $RAR\gamma_{gene\ 1}$ (327 bp SstI-PstI fragment) led to the isolation of $RAR\gamma_{gene\ 3}$. For sequencing, suitable restriction fragments of both the cDNA and the genomic clones were subcloned in Bluescript and double-stranded cDNA was sequenced using the dideoxynucleotide chain-termination method (25) with Sequenase (United States Biochemical).

Transient transfection of CV-1 cells

CV-1 cells were seeded at a density of 1.3×10^6 per plate in DMEM supplemented with 10% fetal calf serum. The following day, cells were transfected with 1μg of expression plasmids and 4 μg of CAT reporter plasmids using a modified calcium phosphate precipitation procedure (23,26). As internal transfection standard, 3μg of β-galactosidase gene (pCH110, Pharmacia) were cotransfected. CAT and β-galactosidase activity were determined as described previously (23).

Primer extension analysis

The end-labeled oligonucleotide primer (5′-TTAAAAGCGAA-CGCTGGAGGGGT-G-3′ annealing to the region +317 to +341 of exon Id) was extended using 50 U MoMLV reverse transcriptase (BRL), cold deoxynucleotides and 50 μg of total RNA derived from the human lung cell line WI-38 (American Type Culture Collection, Ma) as template. The resultant cDNA was analyzed on a sequencing gel.

RESULTS AND DISCUSSION

Isolation of RARγ cDNA clones and functional analysis

To allow structural analysis of the RARγ receptor, cDNA clones were isolated from human cDNA libraries including HL-60 cells, primary keratinocytes, and placental tissue. The longest cDNA clone, RARg1, was isolated from a placental λgt11 expression library using a random primed ^{32}P-labeled cDNA comprising the ligand binding domain of RARα as the hybridizing probe. Sequence analysis revealed that this clone encodes RARγ1 and consists of a 146 bp 5′-untranslated region, an open reading frame of 1362 bp and an 3′-untranslated region of 1064 bp (Figure 1).

```
   1 GCGG...[INTRON]....TCTCTTTCTCCCTAGGACTTTTGGAGGCCCAGTGGGCAGGCCAGGCAGGGCGGGTACGGAGCCTCCCAGGCTGGGGCAGTGGGCATGGGCAGGGGCTGTGGCT
                                                                  M  A  T  N  K  E  R  L  F  A  A  G  A  L  G  P  G  S  G  Y  P  G   22
  93 GAAGACCTCGCCCGCCCACTGCAGACCCCAGGGGACTCTCACACCGCAGCTGCCATGGCCACCAATAAGGAGCGACTCTTTGCGGCTGGTGCCCTGGGGCCTGGATCTGGCTACCCAGGG
     A  G  F  P  F  A  F  P  G  A  L  R  G  S  P  P  F  E  M  L  S  P  S  F  R  G  L  G  Q  P  D  L  P  K  E  M  A  S  L      61
 213 GCAGGTTTCCCCTTCGCCTTCCCAGGGGCACTCAGGGGGTCTCCGCCTTTCGAGATGCTGAGCCCTAGCTTCCGGGGCCTGGGCCAGCCTGACCTCCCCAAGGAGATGGCCTCTCTGTGT
                                                                  S  V  E  T  Q  S  T  S  S  E  E  M  V  P  S  S  P  S  P  P  P  P  R  V  Y  K   88
 331 GAGTCTCAGGGATCT..[INTRON].CTGCTTCCCTCCCAGCGGTGGAGACACAGAGCACCAGCTCAGAGGAGATGGTGCCCAGCTCGCCCTCGCCCCCTCCGCCTCCTCGGGTCTACAAG
     P  C  F  V  C  N  D  K  S  S  G  Y  H  Y  G  V  S  S  C  E  G  C  K                                                  G  F  113
 412 CCATGCTTCGTGTGCAATGACAAGTCCTCTGGCTACCACTATGGGGTCAGCTCTTGTGAAGGCTGCAAGGTGTGTACGTGGTGGG....[INTRON]....TTTCCTCAACTCCAGGGCTTC
     F  R  R  S  I  Q  K  N  M  V  Y  T  C  H  R  D  K  N  C  I  I  N  K  V  T  R  N  R  C  Q  Y  C  R  L  Q  K  C  F  E  V  153
 486 TTTCGCCGAAGCATCCAGAAGAACATGGTGTACACGTGTCACCGCGACAAAAACTGTATCATCAACAAGGTGACCAGGAATCGCTGCCAGTACTGCCGGCTACAGAAGTGCTTCGAAGTG
     G  M  S  K  E                                                  A  V  R  N  D  R  N  K  K  K  K  E  V  K  E  E  G  S  P  D  178
 606 GGCATGTCCAAGGAAGGTAGGCTGCCATCCTT....[INTRON]....GCCTCCACCCACAAGCTGTGCGAAATGACCGGAACAAGAAGAAGAAAAGAGGTGAAGGAAGAAGGGTCACCTGAC
     S  Y  E  L  S  P  Q  L  E  E  L  I  T  K  V  S  K  A  H  Q  E  T  F  P  S  L  C  Q  L  G  K  Y  T  T                  212
 681 AGCTATGAGCTGAGCCCTCAGTTAGAAGAGCTCATCACCAAGGTCAGCAAAGCCCATCAGGAGACTTTCCCCTCGCTCTGCCAGCTGGGCAAGTATACCACGGTGAGGAATGGGC.....
                                N  S  S  A  D  H  R  V  Q  L  D  L  G  L  W  D  K  F  S  E  L  A  T  K  C  I  I  K  I  V  E  243
 783 ..[INTRON]....CCCTCCCCTCTGCAGAACTCCAGTGCAGACCACCGCGTGCAGCTGGATCTGGGGCTGTGGGACAAGTTCAGTGAGCTGGCTACCAAGTGCATCATCAAGATCGTGGAG
     F  A  K  R  L  P  G  F  T  G  L  S  I  A  D  Q  I  T  L  L  K  A  A  C  L  D  I  L                                  271
 876 TTTGCCAAGCGGTTGCCTGGCTTTACAGGGCTCAGCATTGCTGACCAGATCACTCTGCTCAAAGCTGCCTGCCTAGATATCCTGGTAAGTTAGGCCC....[INTRON]....ACCCCCCAC
           M  L  R  I  C  T  R  Y  T  P  E  Q  D  T  M  T  F  S  D  G  L  T  L  N  R  T  Q  M  H  N  A  G  F  G  P  L  T  D  309
 960 CTCCAGATGCTGCGTATCTGCACAAGGTACACCCCAGAGCAGGACACCATGACCTTCTCCGACGGGCTGACCCTGAACCGGACCCAGATGCACAATGCCGGCTTCGGGCCCCTCACAGAC
     L  V  F  A  F  A  G  Q  L  L  P  L  E  M  D  D  T  E  T  G  L  L  S  A  I  C  L  I  C  G                              339
1074 CTTGTCTTTGCCTTTGCTGGGCAGCTCCTGCCCCCTGGAGATGGATGACACCGAGACAGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGGTGCGGGGGGCGCCCCT....[INTRON]..
                     D  R  M  D  L  E  E  P  E  K  V  D  K  L  Q  E  P  L  L  E  A  L  R  L  Y  A  R  R  R  R  P  S  Q  P  Y  374
1165  CTTGTCCCCATAGAGACCGCATGGACCTGGAGGAGCCCGAAAAAGTGGACAAGCTGCAGGAGCCACTGCTGGAAGCCCTGAGGCTGTACGCCCGGCGCCGGCGGCCCAGCCAGCCCTAC
     M  F  P  R  M  L  M  K  I  T  D  L  R  G  I  S  T  K                                                  G  A  E  R  A  I  T  399
1269 ATGTTCCCAAGGATGCTAATGAAAATCACCGACCTCCGGGGCATCAGCACTAAGGGTGAGTCATGAGTGG....[INTRON]....TCTCTCTCCATTGTAGGAGCTGAAAGGGCCATTACT
     L  K  M  E  I  P  G  P  M  P  P  L  I  R  E  M  L  E  N  P  E  M  F  E  D  D  S  S  Q  P  G  P  H  P  N  A  S  S  E  D  439
1344 CTGAAGATGGAGATTCCAGGCCCGATGCCTCCCTTAATCCGAGAGATGCTGGAGAACCCTGAAATGTTTGAGGATGACTCCTCGCAGCCTGGTCCCCACCCCAATGCCTCTAGCGAGGAT
     E  V  P  G  G  Q  G  K  G  G  L  K  S  P  A  END                                                                        454
1464 GAGGTTCCTGGGGGCCAGGGCAAAGGGGGCCTGAAGTCCCCAGCCTGACCAGGGCCCCTGACCTCCCCGCTGTGGGGGTTGGGGCTTCAGGCAGCAGACTGACCATCTCCCAGACCGCCA
1584 GTGACTGGGGGAGGACCTGCTCTGCCCCTCTCCCCACCCCTTCCAATGAGCTCCTTGTTTTTGCCAAAGTTTCTAGGGGTGCCTCTGTGTTCATCCCCTTCCTGATCTAACCGGCTCCCTC
1704 GCCAGTCCCGGGGGGCCTGCCCTGCTCCCACCAGGAGAGAGGGCAAAGGGATGAGCCTGGGTTTGGACTCTAAAATCTCAGCACTGCCCCATGGGTCCTAGACTTCCCAGGGCAAGAGGAA
1824 GACCCTGCCATTCCACAGCCCCTTCCTCTGCCAGGTGCTTGGCTCTCTGAGAGCAAACAGGAACACTAGAGACCAAAAAGGGCACAAAGGAGAAGGGTGAGCCCACCTTCTTGCTCCTAC
1944 CCTTGGTGCCTAATGCTGTGTGATGCACCTGCAGGGTGTGTGCTAGCCTCTGTGCCCCGTCCTTGTGCCAGGTCAAGGTGGGGGCAAGCTGGGCCCTGCATTTCTGGGGCAGGAACAGAG
2064 GGTGAAAGGGACAGATAGATGCAGGTCCATTCTGCACCTCTTGGCTCGGGTGCAGAGTTCACCCTGTGCCCTCCGTTATAAGTCCCTCCCCCAGCCCTGTCATGTGCCTTGGGCTCCTCC
2184 TGCCCTCCATCTCAGCCATTGGGGCAGGGACCCTCCTACACTACAGAGGGGCCAGGGGATCCCTCTCTCCCTAGTGCCTTCCACCCTTTACTCCCCCAGAGCAGCTTGGCCCAGGGAGGG
2304 GGGATGCTGCTTAGCTGATCCCGCCCTGACCCAGAGGAAGCCTCTATTTATTTATTAGCTTTTGTTTACACCGTGGAATTGACCCTTCCTCCAGGGGTCTTGGGTGGGGGAGCCCAGGGC
2424 CCCTGTGACCCCTCCTTTCTTCCTCCAATCCCCAGTTTGTATTTAGCTGCCAAATAAGATTCCCCATTGGCTCCCTGTGTTCTCTTGGGGGGTCAGGGTGCTGTCCCCTCCCCTCTGTTT
2544 ACATCTCCCCTCTACCCCGCTGTATCGC
```

Figure 1. Nucleotide and deduced amino acid sequence of the RARγ1 cDNA including flanking intron sequences. The consensus polyadenylation signal AATAAA at the position 2475 is underlined. The selective mRNA degradation signals ATTTA in the 3′-untranslated region are shown in bold letters. The canonical donor and acceptor splice sites are underlined.

This contains the consensus polyadenylation signal AATAAA at position 2475 (on the genomic level a second AATAAA sequence located 123 bp further downstream was identified). The selective mRNA degradation signal ATTTA occurs three times in the 3'-untranslated region suggesting a short half-life of RARγ mRNA (27). In comparison to the sequence of RARγ published by Krust et al. (13), the RARγ1 cDNA sequence is 938 bp longer in the 3'-untranslated regions, reveals one silent point mutation (Pro-75 [CCA−CCC]), and did not contain the nucleotide A-1888 (numbering according to Krust et al., [13]). These findings were confirmed later by the genomic sequence.

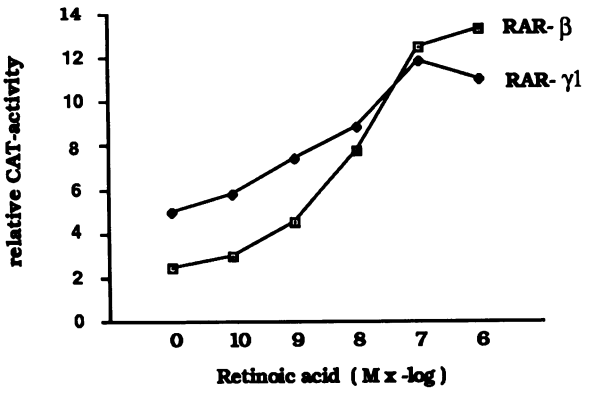

Figure 2. Retinoic acid dose dependent activation of the TRE_2-*tk* CAT reporter construct by RARβ and RARγ1. CV-1 cells were cotransfected with 4 μg reporter plasmid and 1 μg of the receptor expression vectors. Cells were treated with no hormone or increasing amounts of RA (100 ϱM-1μM) and assayed for CAT activity 24 hrs later. Relative CAT activity was normalized for β-galactosidase activity. A representative experiment is shown.

To analyse the biological activity of the encoded RARγ protein, we inserted the coding sequence into a eukaryotic expression vector (pECE) and cotransfected pECE-RARγ together with a RA-responsive reporter gene (TRE_2-*tk* CAT) into CV-1 cells. TRE_2-*tk* CAT contains two RA-responsive elements upstream of the HSV-*tk* gene promoter pBLCAT2 construct (28,29). We observed that addition of RA resulted in a concentration dependent activation of the reporter gene (Figure 2). RARγ1 shows higher basal level activity than RARβ, shown here as a control, and displayed a more sensitive response to RA than RARβ. Our data clearly demonstrates that the isolated cDNA clone encodes an active RARγ1 protein.

Structural organization of the RARγ gene

Two overlapping clones ($RAR\gamma_{gene1-2}$) were isolated from a human genomic placental λ FIX II library using RARγ cDNA as a ^{32}P-labeled hybridizing probe. The clones span a region of 21.5 kb and represent the entire open reading frame of the RARγ

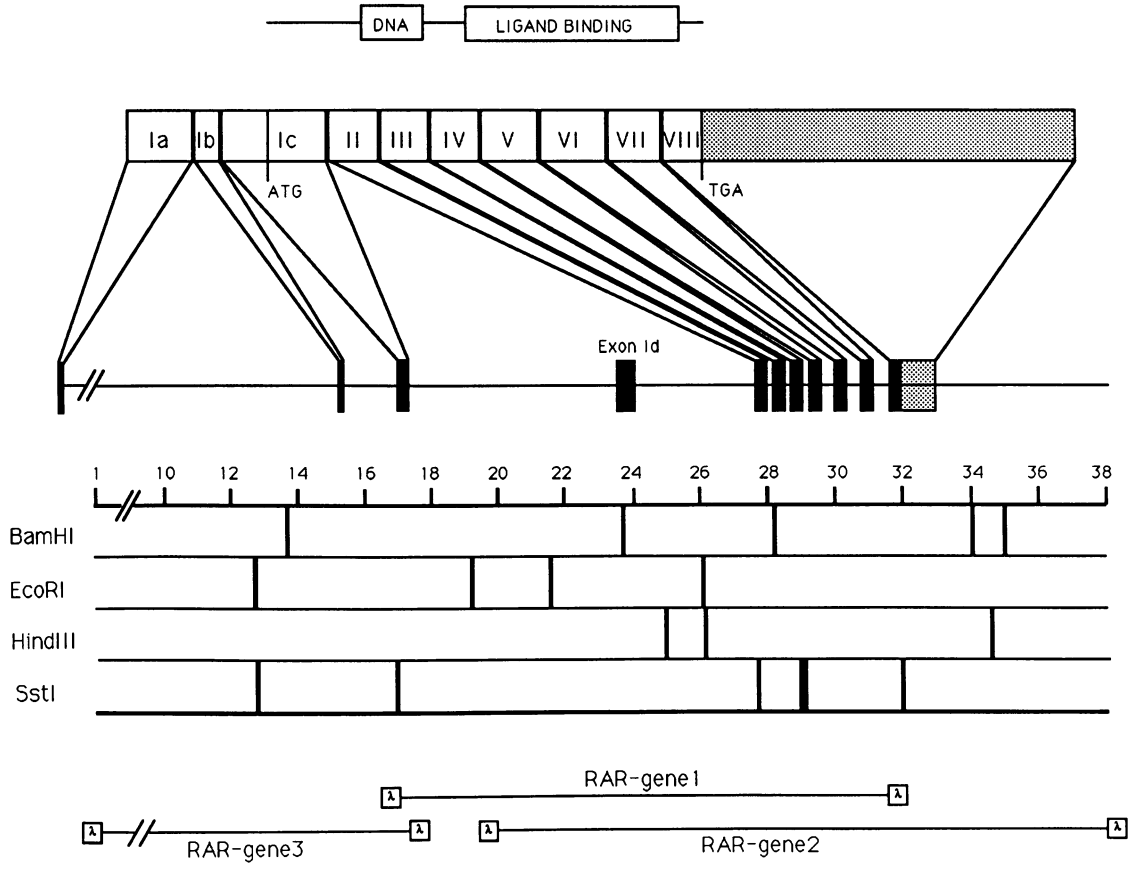

Figure 3. Organization of the human RARγ gene locus. The ten exons encoding the human RARγ1 (solid boxes) as well as the exon encoding the alternative N-terminal region of RARγ2 are shown on the genomic DNA. The corresponding positions of the exons with respect to the RARg1 cDNA clone are shown above. Numbers at the borders refer to the cDNA nucleotide sequence (Figure 1). Position of the translation initiation and termination codons are indicated. At the top, the location of the DNA and the ligand binding domains are shown. A restriction map of the RARγ gene covered by the genomic clones $RAR\gamma_{gene1-3}$ and their orientation are presented at the bottom.

gene. A third clone ($RAR\gamma_{gene3}$) was isolated using a probe derived from the 5'- end of clone $RAR\gamma_{gene1}$. Figure 3 shows the restriction map covering approximately 38 kb of genomic sequence. To delineate the exon-intron organization, the coding exons and flanking intron regions were sequenced (Figure 1). Intron/exon boundaries have the canonical consensus splice sequences. The complete gene is composed of eleven exons, of which ten are used for the RARγ1 transcript (Figure 3, exons Ia, Ib, Ic, and II–VIII). The most 5'- exons of RARγ1 is not contained in the 38 kb sequence, but is predicted from the cDNA sequence (13). An additional exon (Exon Id) defining the RARγ2 5'- message was detected within intron 3 of the gene (Figure 3). Seven exons—common to all RARγ isoforms—are containing the complete RARγ coding sequence with the exception of the N-terminal domain, and were located within approximately 5 kb of the 3'-end of the RARγ gene. This rather compact portion of the gene may represent an ancestral version of the RARγ gene since an RARγ1 protein deleted for the A-region has transactivation properties similar to RARγ2 (data not shown).

Exon structures common for RARγ isoforms

From the eleven RARγ coding exons, seven (exons II to VIII) are common to all RARγ isoforms. They encode domains B through F and the entire 3'-untranslated region as shown in Figure 3. Exon II (Ser-62 to Lys-111) encodes the entire B-region together with the first 'zinc-finger' of the DNA-binding domain. This is in contrast to steroid receptors which encode the large A-region together with most of the B-region in a single exon (30–33). The splice site between Exon II and III (i.e. Lys-111) is identical to that found in RARβ and chicken thyroid hormone receptor TRβ (34). This location is characteristic for the thyroid/retinoic acid receptor subfamily and differs from the also highly conserved position found in genes of the steroid receptor subfamily (i.e. 8 amino acids downstream; 30–33). Ponglikitmongkol et al. (32) have suggested the existence of an ancestral nuclear receptor gene that contained both introns. During the evolutionary divergence of the two subfamilies one intron was selectively lost. However, the situation seems to be more complicated, since the N10 gene, encoding a nuclear

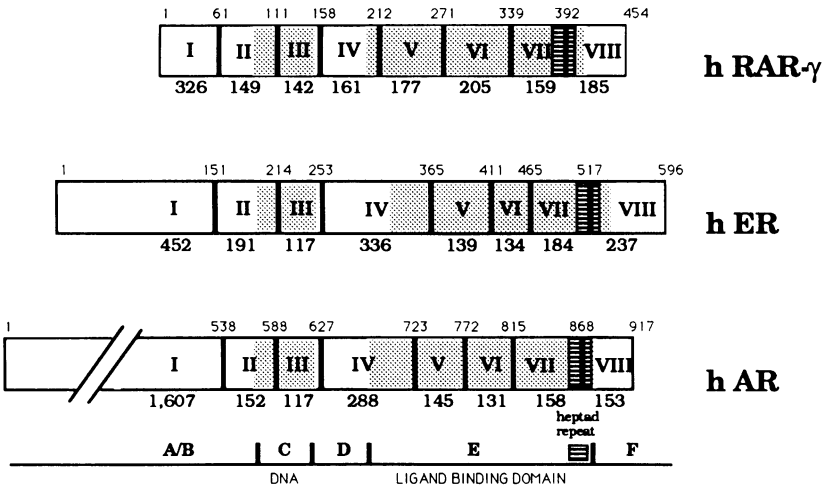

Figure 4. Comparison of the exon-intron structure of RARγ1, with human estrogen receptor (hER, 32) and human androgen receptor (hAR, 31). Splice sites are indicated by bold vertical lines and the number of the amino acid residue. DNA and ligand binding domains are shown as shadowed boxes. The locations of the regions A-F and of the heptad repeat are shown in the bottom line.

```
-1219                                                    CTGCAGCAGGGGAGTGGCCCCCCACCCGGTACAACAGATTCTTAGAAAAGCGCGCCCCGCCCTCCGGGAGCCACAGC
-1142 CGCCTGCGGGAAGGGCCGGGCCGGGCGCCGCTCCCGCTCCGCCGCTGCCCGTGTGAGCGGCGCTGACAGAGCCGGCTGGGCAGGCAGGCCGGGCGCCGCGGGCTGAGGGCGGGGCGCGGG
-1022 GGAGGGGGGGATTTCCGGGAGTCGCCGCTGCTGGCGCTGTCCGGGGCCAGGGTCCCCCCATTCGGCCCTGCCGGGCCAGGGCAAACTTCTGACGTGGCCAGCCCCCCGCCCACCTCACCC
 -902 CCTTCTCCGCCGGGCGAGGGCGCTCAGTGCCGGAAGGGGCAACCCAGGGAACAGTGTTGCGGCCTTGGAGAGCTGAGGGGCACTAACCGAAGGGAGGGCAGTTATGCCTAGGCTAGAGGG
 -782 TATTTAAAGACCCTGAGAACAGTCAGTCAGTGACCAGAAGCGCCCCTATTGCCGAGGGAGCCCTTGGCAGTGAGAAACTGGAGGAGGGGAGTTGGCCATCAGGGGCAGATTCATCGAGGC
 -662 GCAGTTGGCACGGCAGGCACAGCTTATGGGGTAGTTTTGTGCTCTAAAAGTGAAGTTATGTTCGAAGAAGGAGCATCGAGAGAGACGGTCGAGGCGCTTAGGGGGCGTCAGTTGCGGGAT
 -542 AAGGCTTCGGGGACTCGGAGGGAGGCGGCCCGGGCCGCGGCGGAGGGCGGACTGCTCGGGCTAGGAGGGGTAGACTCTGGCTGCGCTGGAGGGGGCGGGGGGACTCCGCGGCAGCTGAAG
 -422 ACGGCCTGGGCGGGGTGGGGCGGGGCCGGGTCAGGGAGGAGGGTGAGCGCGCTCGGCCACCCGCCCCGAGGGGTATCCCAGAGAGTTCACGGCGGGACGCGCTCCGGCCCCTAGGACCGC
 -302 GGGTGGGGAGACGGGAAGGAAGGAGGAGGCACGCGCGCCGAGCCGCAGGAATGCACTTTTCGGCCCTGAAGCGGGCGCGCGGAGGGCGAGGGCGGAGCCAGGCCTGGGAATGAGAGGGGT
 -182 GGAGGTAGTGCAGGTCCCGAGCCTCAGGAGAGCAGCTGCGCGCGGGGACCTCCACGTTCCCCTCTCCCCAGTCATCCTTAACACCCCTGTATTGACTAGTGTGAAAAAGGGAGAGTTTGG
  -62 GGAGAGGGAAAACTCTGAAAGGAAATGCTTATGCTTGGTTCCTCCTCTCCTTCTTCCAACCCCGACCTGGGCACGCGCCTCAGTGCAAATGCATGCAAAGCCTATGAGCTGTGGAGAGTA
   59 TGCCTGCAAGCATTCTCCCGAGCAAGGTGTGTTTGCACAAACGTGTGTGCACCTATGATTTATATGCGTGCTCACATGTATGTTCACTCTTGCATATTTGCAAATTCTTGGTGCCTGATA
  179 AGATAAATGCATGCACGGATGTGTGTCTAAGCAAGCGCGCGCGCGGAGCGAGTGTACGTATGTGTATGTGTGCGCTCCACTCTCAGCTTTGCTCATCTCGGTGCCCGGCCCCCTTTGC
  299 TCTCCAAGCCCCACCCCCCACCCCTCCAGCGTTCGCTTTTAAAAAAGGTCTCCCCAGTGCTAGCTGCCGAAGCACCCAGATAAGAGCTGGACGGAGGCTAAAGCGGCAGCCCCACGTTCG
  419 CGCCCCGCCCAGTCCCTTTCCCCTGCTGGGATCCCCCCTCTTCCCCGCCCACCCCTTACCCGCATGCAGCCCAGCGCCCTATGCTAGCCCTCCCCCTCCCCCCCTGCTGGAGCGGGGCCC
                                                                                                                    M Y D C M E T F A P G P R R L   15
  539 GCCGGGGGAGGAGGGGGAATCGGCTGCGGGTCCTTGGTGTTTCCAGCACCCAGTTTCCCTTCAAGCCGGGTCGCGATGTACGACTGTATGGAAACGTTTGCCCCGGGTCCGCGACGGCTG
    Y G A A G P G A G L L R R A T G G S C F A G L E S F A W P Q P A S L Q  ...INTRON   50
  659 TACGGGGCGGCCGGGCCCGGGGCCGGCTTGCTGCGCAGAGCCACGGGCGGCTCCTGTTTCGCCGGACTTGAATCTTTTGCCTGGCCGCAACCCGCCAGCCTGCAAGTAGGTACCGGAGTC
```

Figure 5. Nucleotide and deduced amino acid sequence of exon Id and 5'-flanking intron sequences. The canonical donor splice site is underlined. The asterisks indicate the two major transcription start points. The numbering of the nucleotide sequence on the left refers to the first transcription start point.

receptor with high sequence homology to the nuclear receptor superfamily, contains a third alternative position for the splice site (35). For the glucocorticoid receptor a region designated the P-box, located at the base of the first finger, seems to be critical for the identification of cognate DNA sequence. A second element, the D-box, is involved in the determination of the spacing between the response element half-sites (36–38). Similar P- and D-box motifs were found in all nuclear receptors. Interestingly the P-box motifs of RARs,TRs, and the estrogen receptor are identical (i.e. EGCKG, single letter amino acid code) which may explain the binding of all three receptor types to overlapping sets of response elements (i.e. TRE_{pal} and RARE β-gene-promoter; 17,28,29,38). Since both 'zinc-fingers' seem to have different functions and their amino acid sequence is also highly divergent outside the cysteines, it is possible that both exons have developed separately rather than by exon duplication.

Exon IV encodes the hinge region located between the DNA binding and the RA binding domain. The splice site between exons III and IV is located within a region conserved in all nuclear receptor genes known so far.

The hormone binding domain is encoded by the exons V (Asn-213 to Leu-271), VI (Met-272 to Gly-339), VII(Asp-340 to Lys-391), and by parts of exon VIII(Gly-392 to Met-421). The remainder of exon VIII encodes for the rest of the molecule and the entire 3'-untranslated region. This exon-intron structure of the RARγ gene in the hormone binding domain is different from that found in steroid receptor genes, except for the location of the splice junction within the heptad repeat (Figure 4). The heptad repeat (located between amino acids 373–407 in RARγ) has been proposed to mediate the receptor dimerization function and is distributed on the exons VII and VIII. Fawell et al. (10) have recently observed that the N-terminal portion of this heptad repeat, namely that encoded by exon VII forms part of the dimer interface, while the C-terminal part may contribute to the generation and/or stability of the estrogen receptor-dimer structure. The RA binding function has not yet been assigned to a specific region within the large hydrophobic ligand binding domain. Mutational analysis performed in the estrogen receptor indicates that residues 518 to 525 (i.e. overlapping with the heptad repeat) are essential for hormone binding (10).

Intron exon structures for distinct RARγ isoforms

Exon Ic encodes 142 bp of the 5'-untranslated region and the first 61 amino acids. This exon is unique for RARγ1 and distinguishes this isoform from RARγ2 (13). We have recently found that this part of the N-terminal region is not necessary for the gene activation functions of RARγ1, since a protein lacking this region is still a potent activator of RARE containing promoters (unpublished data, Husmann, JML, MP). The mRNAs encoding RAR isoforms lacking this sequence also appear to be produced in certain tissues (18). It is therefore possible that the exon Ic is a relatively late addition to the RARγ gene. RARγ1 contains at least two additional exons for 5'-untranslated regions (Figure 3), one of which, exon Ia, is not included on the 38 kb analyzed here, and exon Ib which encodes further 67 bp of the 5'-untranslated region. Thus the RARγ1 promoter region is located far upstream of the major RARγ coding sequences.

As mentioned above two distinct RARγ isoforms have been isolated recently. RARγ1 and RARγ2 differ completely in their N-terminal regions while the remainder of the molecule is identical, suggesting that they are derived from a single gene locus. Using restriction analysis and RARγ2 specific primers, the region encoding the A-region of RARγ2 was identified within the 11.0 kb intron located between exons Ic and II. This exon, designated exon Id, is located about 7 kb downstream of exon Ic and encodes 50 amino acids (Figure 5). Amino acid sequence comparison showed a considerable homology with the A-region of RARβ which is most striking around two cysteines and two prolines, suggesting a sandwich-like structure of β-sheets (Figure 6). The functional significance of this (i.e. possible interaction with common transcription factors) remains to be analyzed. Analysis of genomic sequences upstream to sequences homologous with the longest human cDNA clones described, revealed a high sequence homology (83%) up to the position +328 with the entire 5'- untranslated region described for mouse cDNA clones (18). This excludes the presence of intronic sequences. The high conservation of the 5'- untranslated region may also indicate a functional importance of this region. Primer extension analysis using total RNA derived from WI-38 human lung cells as template, defines two major transcription start points (+1 and +5, Figure 7). This indicates that alternative promoter usage rather than alternative splicing is directing RARγ2 synthesis. The sequence motif TTATG at position −34 may serve as a weak TATA-box. The close proximity of the RARγ2 amino-terminal coding exon to the RARγ common coding sequences and the homology observed in the amino acid sequence with RARβ suggest an earlier connection of this exon with RARγ than with the RARγ1 specific exons.

CONCLUSION

Retinoic acid and the thyroid hormones exert profound functions in such complex processes as tissue development and differentiation. The nuclear receptors binding these molecules

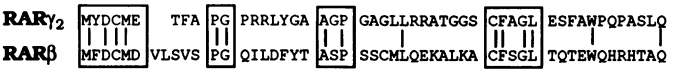

Figure 6. Sequence homology between the N-terminal A-regions of RARγ2 and RARβ. Lines indicate identical amino acids. Regions of sequence similarity are boxed.

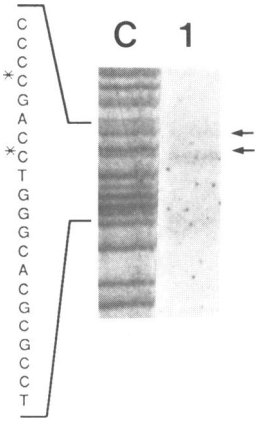

Figure 7. Analysis of the transcription start point of the RARγ2 by primer extension analysis. An end-labeled primer (see materials and methods) was extended by using total RNA from human WI-38 lung cells as template (lane 1). For size verification a sequencing reaction terminated with ddCTP and using the same primer is shown (lane C). The sequence flanking the transcription start point is shown. The asterisks indicate the two major transciption start points.

are structurally highly related and form a subfamily within the steroid/thyroid hormone receptor superfamily. These receptors combine various functional properties such as sequence specific DNA binding, ligand binding, ligand dependent and independent *trans*-activation, receptor-dimerization, and nuclear localization. Genomic analysis revealed that the exon-intron structure of the RARγs is distinct from that found in the steroid receptors (i.e. differences in the length of exons and position of various splice sites). This suggests that during evolution, the TR/RAR subfamily has developed independently from the other nuclear receptors. Noteworthy is that the organization of nuclear receptors into A-F domains based on sequence homology is not strictly consistent with the exon-intron structure. A second important difference within the RAR/TR subfamily is their capability to generate variants with distinct amino- or carboxy-terminal regions. Northern blot analysis revealed spatio-temporal differences in the expression of these variant receptors during embryogenesis (16,18). Based on the structural similarity of the DNA binding domain, TRs and RARs can bind to common response elements, however *trans*-activation may need interactions with other tissue specific transcription factors. Interactions may be elicited by the different N-terminal regions which have been shown to confer tissue specific *trans*-activation function in other nuclear receptors (33,39,40). The compact organization of the major portion of the RARγ coding sequence compared to the diffuse organization of the amino-terminal coding sequences suggest independent origins for these two portions of the RARγ gene. The compact RARγ basic structure may represent the ancient RARγ gene where the large 5'-region of the gene may be a relatively new addition that allows increased flexibility for RARγ regulation and isoform expression using multiple promoters.

ACKNOWLEDGMENTS

We thank W.Lernhardt for providing the genomic library. We also thank E.D.Adamson and R.A.Maki for critical reading the manuscript, and Kathleen Sweeting for secretarial assistance. J.M.L. and B.H. were supported by a fellowship from the Deutsche Forschungsgemeinschaft. This work was supported by NIH grant DK35083 and NCI grant CA50676.

REFERENCES

1. Roberts, A.B. and Sporn, M.B. (1984) *The Retinoids*, **2**, 209–286.
2. Thaller, C., and Eichele, G. (1987) *Nature*, **327,** 625–628.
3. Thaller, C., and Eichele, G. (1990) *Nature*, **345,** 815–819.
4. Giguere, V., Ong, E.S., Segui, P., and R.M. Evans. (1987) *Nature*, **330,** 624–629.
5. Petkovich, M., Brand, N.J., Krust, A., and P. Chambon. (1987) *Nature*, **330,** 444–450.
6. Benbrook, D., Lernhardt, E., and M. Pfahl. (1988) *Nature*, **333,** 669–672.
7. Brand, N., Petkovich, M., Krust, A., de The, H., Marchio, A., Tiollais, P., and A. Dejean. (1988) *Nature*, **332,** 850–853.
8. R.M. Evans. (1988) *Science,* 240, 889–895.
9. Green, S., and P. Chambon. (1988) *Trends Genet.*, **4,** 309–314.
10. Fawell, S.E. Less, J.A., White, R., and M.G. Parker. (1990) *Cell,* 60, 953–962.
11. Forman, B.M., and H.H. Samuels. (1990) The New Biologist, **2**, 587–594.
12. Zhang, X-k, Wills,K.N., Graupner, G., Tzukerman, M., Hermann, T., and M. Pfahl, (1991) The New Biologist, (in press).
13. Krust, A., Kastner, P.H., Petkovich, M., Zelent, A., and P. Chambon. (1989) *Proc. Natl. Acad. Sci. USA*, **86,** 5310–5314.
14. Giguere, V., Shago, M., Zirngibl, R., Tate, P., Rossant, J., and S. Varmuza. (1990) *Mol. Cell. Biol.*, **10,** 2335–2340.
15. Dolle, P., Ruberte, E., Kastner, P., Petkovich, M., Stoner, C.M., Gudas, L.J., and P. Chambon. (1989) *Nature*, **342,** 702–704.
16. Ruberte, E., Dolle, P., Krust, A., Zelent, A., Morriss-Kay, G., and P. Chambon. (1990) *Development*, **108,** 213–222.
17. Hoffmann, B., Lehmann, J.M., Zhang, X.-K., Hermann, T., Husmann, M., Graupner, G., and Pfahl, M. (1990) *J. Mol. Endo.*, **4,** 1734–1743.
18. Kastner, Ph., Krust, A., Mendelsohn, C., Garnier, J.M., Zelent, A., Leroy, P., Staub, A., and P. Chambon. (1990) *Proc. Natl. Acad. Sci. USA*, **87,** 2700–2704.
19. Hodin, R.A., Lazar, M.A., Wintman, B.I., Darling, D.S., Koenig, R.J., Larsen, P.R., Moore, D.D., and W.W. Chin. (1989) *Science*, **244,** 76–78.
20. Forrest, D., Sjoberg, M., and B. Vennstrom. (1990) *EMBO J.*, **9,** 1519–1528.
21. Benbrook, D., and M. Pfahl. (1987) *Science*, **238,** 788–791.
22. Millan, J.L. (1986) *J. Biol. Chem.*, **216,** 3112–3115.
23. Pfahl, M., Tzukerman, M., Zhang, X-K., Lehmann, J.M., Hermann, T., Wills, K.N., and G. Graupner. (1990) *Methods Enzymol.*, **189**, 256–270.
24. Maniatis, T., Fritsch, E.F., and J. Sambrook. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. Sanger, F., Nicklen, S., and A.R. Coulson. (1977) *Proc. Natl. Acad. Sci. USA,* 74, 5463–5467.
26. Chen, C.A., and H. Okayama. (1988) *BioTechniques*, **6,** 632–637.
27. Shaw, G., and R. Kamen. (1986) *Cell*, **46,** 659–667.
28. Graupner, G., Wills, K.N., Tzukerman, M., Zhang, X-K., and M. Pfahl. (1989) *Nature*, **340,** 653–656.
29. Graupner G., Zhang, X.-K., Tzukerman, M., Wills, K.N., Hermann, T. and Pfahl, M. (1991) *J. Mol. Endo.* (in press).
30. Jeltsch, J.M., Turcotte, B., Garnier, J.M. Lerounge, T., Krozowski, Z., Gronemeyer, H., and P. Chambon. (1990) *J. Biol. Chem.*, 265, 3967–3974.
31. Lubahn, D.B., Brown, T.R., Simental, J.A., Higgs, H.N., Migeon, C.J., Wilson, E.M., and F.S. French. (1989) *Proc. Natl. Acad. Sci. USA*, **86,** 9534–9538.
32. Ponglikitmongkol, M., Green, S., and P.Chambon. (1988) *EMBO J.*, **7,** 3385–3388.
33. Gronemeyer, H., Turcotte, B., Quirin-Stricker, C., Bocquel, M.T., Meyer, M.E., Krozowski, Z., Jeltsch, J.M., Lerouge, T., Garnier, J.M., and P. Chambon. (1987) *EMBO J.*, **6,** 3985–3994.
34. Zahraoui, A., and G. Cuny. (1987) *Eur.J.Biochem.*, **166,** 63–69.
35. Ryseck, R.P., Macdonald-Bravo, H., Mattei, M.G., Ruppert, S., and R. Bravo. (1989) *EMBO J.*, **8,** 3327–3335.
36. Mader, S., Kumar, V., de Verneuil, H., and P. Chambon. (1989) *Nature,* 338, 271–274.
37. Umesono, K., and R.M. Evans. (1989) *Cell*, **57,** 1139–1146.
38. Glass, C.K., Lipkin, S.M., Devary, O.V., and M.G. Rosenfeld. (1988) *Cell*, **54,** 313–323.
39. Miesfeld, R., Godowski, P.J., Maler, B., and K.R. Yamamoto. (1987) *Science*, **236,** 423–427.
40. Hollenberg, S.M., Giguere, V., Segui, P., and R.M. Evans. (1987) *Cell*, **49,**39–46.